

RESEARCH ARTICLE

Prediksi Diagnosis Hepatitis B Virus Menggunakan *Gated Graph Neural Network*

Fadhil Wisnu Ramadhan and Kemas Rahmat Saleh Wiharja*

Fakultas Informatika, Universitas Telkom, Bandung, 40257, Jawa Barat, Indonesia

* Corresponding author: bagindokemas@telkomuniversity.ac.id

Abstrak

Hepatitis merupakan infeksi virus pada hati dan dapat menyebabkan komplikasi terhadap penyakit lain yang dialami oleh pasien. Diagnosis dini dan penanganan yang tepat sangat penting untuk mencegah progresi penyakit dan komplikasi lebih lanjut. Diperlukan sebuah sistem prediksi diagnosis hepatitis yang akurat untuk menangani dan mengatasi kemungkinan terjangkitnya seseorang akan hepatitis. Penelitian ini melakukan prediksi model *Gated Graph Neural Network* terhadap pilihan data *Hepatitis UCI Machine Learning Repository*. Pada penelitian ini dilakukan pemodelan dan penelitian model dengan dua model *graph neural network* lainnya dan menghasilkan evaluasi yang baik pada prediksi klasifikasi *node* hepatitis, dengan menggunakan *Gated Graph Neural Network* model menunjukkan nilai yang superior terhadap 2 metode lain yaitu *GAT* dan *GCN*. Di mana *GGNN* mendapatkan nilai *Accuracy*, *Precision*, dan *Recall* di atas 90%.

Key words: Hepatitis, *Gated Graph Neural Network*, Prediksi.

Pendahuluan

Hepatitis adalah kondisi inflamasi pada hati yang disebabkan oleh berbagai virus infeksi, yang dapat menyebabkan berbagai masalah kesehatan, dengan beberapa kasus yang berpotensi fatal. Terdapat lima jenis utama virus hepatitis, yang dikenal sebagai tipe A, B, C, D, dan E. Meski semuanya menyebabkan penyakit hati, mereka memiliki perbedaan penting dalam hal cara penularan, tingkat keparahan, distribusi geografis, dan metode pencegahan. Secara khusus, tipe B dan C dapat menyebabkan penyakit kronis yang mempengaruhi ratusan juta orang dan menjadi penyebab utama sirosis hati, kanker hati, serta kematian akibat hepatitis virus. Diperkirakan sekitar 354 juta orang di seluruh dunia hidup dengan hepatitis B atau C, dan bagi kebanyakan orang, pengujian dan pengobatan masih belum dapat dijangkau [1]. Prediksi yang akurat tentang perkembangan dan hasil infeksi Hepatitis menjadi hal yang penting untuk pengobatan dan penanganan penyakit. Terutama dengan kemajuan teknologi dalam bidang *Artificial Intelligence* dan *Machine Learning*, sebuah sistem dimana diagnosis yang lebih cepat dan akurat dapat sangat membantu penanganan pasien yang terjangkit Hepatitis. Sudah ada beberapa penelitian sebelumnya yang menggunakan *Machine Learning* dalam melakukan prediksi virus hepatitis, dengan menggunakan klasifikasi pembelajaran mesin yang diantaranya adalah menggunakan *decision tree*, *logistic regression*, *super vector machine*, *random forest*, *Xgboost*, dan juga *AdaBoost*. Obaido et al. [2] melakukan penelitian tersebut menggunakan model model yang dipilih dikarenakan oleh popularitas dan penggunaan reguler di beberapa kasus deteksi penyakit. Popularitas tersebut didapatkan

karena model tersebut dapat melakukan pembelajaran model terhadap dataset dengan dimensi yang banyak dan *imbalance*, mengetahui paper ini menggunakan dataset UCI sebagai data untuk modelnya. Penggunaan metode tersebut ini menunjukkan bahwa model tersebut dapat melakukan performa prediksi yang baik dan efektif terhadap data klinis yang tersedia.

Ibrahim et al. [3] melakukan prediksi model terhadap infeksi hepatitis b dengan menggunakan *artificial neural network*, dan *genetic algorithm*. Dimana data yang digunakan pada penelitian ini diambil dari 900 data medis pada . Beberapa fitur yang digunakan dalam penelitian ini dilakukan dengan mengambil gejala gejala yang dialami oleh pasien. Dari hasil penelitian tersebut dihasilkan bahwa *accuracy* yang didapat pada model menunjukkan akurasi sebesar 66,30%, spesifisitas 66,33%, dan sensitivitas 77,53%. Prediksi tersebut dianggap memiliki ukuran kinerja yang dapat diterima, yang diharapkan dapat membantu para ahli medis merespons lebih cepat di masa depan dan mengurangi wabah yang mungkin terjadi di seluruh dunia. Dengan berkembangnya model *neural network* dalam melakukan prediksi model terhadap suatu penyakit, representasi data dalam bentuk graf mulai mendapatkan perhatian yang cukup baik beberapa tahun ini [4], data graf dapat membantu merepresentasikan data dalam bentuk yang lebih kompleks dengan dimensi yang tinggi, dimana relasi antar *node* dan *edge* dari sebuah data dapat di tentukan sebagaimana kuat relasinya terhadap *node* yang lain. Dengan itu penelitian ini mengajukan sebuah model untuk memprediksi infeksi Hepatitis pada sebuah pasien dengan menggunakan *Gated Graph Neural Network* (GGNN). *Gated graph neural*

networks (GGNN) adalah jenis GNN tertentu yang telah menunjukkan hasil yang menjanjikan dalam memodelkan grafik yang kompleks dan memprediksi hasil dalam berbagai bidang. GGNN menggunakan mekanisme gerbang dimana *gate* dalam GRU berfungsi seperti sebuah filter yang memungkinkan model untuk mempelajari representasi yang lebih kaya dan dinamis dengan menyesuaikan kontribusi dari informasi yang datang pada setiap iterasi. Dengan demikian, GGNN mampu mempertahankan konteks informasi lama sekaligus menyesuaikan informasi baru yang datang, memberikan model kemampuan untuk memproses informasi pada jaringan graf secara lebih efektif dan akurat.

Dalam konteks prediksi infeksi Hepatitis, GGNN dapat digunakan untuk memodelkan interaksi antara pasien dengan hasil tes laboratorium maupun demografinya. Dengan adanya keterkaitan antara hasil tes diagnosis laboratorium seorang, GGNN yang berbentuk graf dapat melakukan prediksi dengan *node* sebagai pasien dan data pasien sebagai fitur *node*.

Topik dan Batasannya

Berdasarkan latar belakang yang telah diurai, maka dapat dirumuskan bahwa permasalahan yang dicakup adalah:

1. Bagaimana implementasi model *Gated Graph Neural Network* dalam melakukan prediksi diagnosis hepatitis?
2. Bagaimana analisis performa akurasi klasifikasi yang dihasilkan oleh model *Gated Graph Neural Network* terhadap dataset Hepatitis UCI?

Berikut adalah ruang lingkup penelitian tugas akhir ini:

1. Dataset yang digunakan memiliki beberapa nilai yang hilang.
2. Data yang digunakan adalah data survey Hepatitis UCI *Machine Learning Repository*, dengan hanya terdiri dari 19 fitur.

Tujuan

Tujuan dari Tugas Akhir ini adalah untuk melakukan implementasi model prediksi diagnosis pasien yang terjangkit Hepatitis dengan menggunakan *Gated Graph Neural Network*, dengan hasil prediksi yang akurat sebagai acuan utama dalam perancangan model.

Tinjauan Pustaka

Setelah dilakukan telaah lebih lanjut terhadap beberapa penelitian, ada beberapa penelitian yang memiliki keterkaitan terhadap penelitian yang akan dilakukan. Sun et al. [5] mengusulkan model GNN dalam prediksi penyakit, yang menggunakan rekam medis elektronik (EMR). Model ini belajar membuat *node embeddings* yang sangat representatif untuk pasien, penyakit, dan gejala dari grafik konsep medis serta grafik rekam medis pasien, yang masing masing dibangun dari basis pengetahuan medis dan EMR. Dengan menggabungkan informasi dari *node* tetangga yang terhubung langsung, *encoder neural graph* yang diusulkan dapat secara efektif menghasilkan *embeddings* yang menangkap pengetahuan dari kedua sumber data, dan mampu secara induktif menyimpulkan *embeddings* untuk pasien baru berdasarkan gejala yang tercatat dalam EMR mereka. Hal ini memungkinkan prediksi yang akurat baik untuk penyakit umum maupun penyakit langka. Percobaan ekstensif pada dataset EMR dunia nyata telah menunjukkan kinerja unggul dari model yang diusulkan.

Obaido et al. [2] mengajukan sebuah model prediksi diagnosis hepatitis b dengan menggunakan dataset UCI, metode yang digunakan dalam penelitian ini adalah *decision tree*, *logistic regression*, *super vector machine*, *random forest*, *adaboost*, dan *xgboost*, dan metode tersebut mencapai akurasi seimbang masing masing sebesar 75%, 82%, 75%, 86%, 92%, 90%. Ibrahim et al. [3], mengerjakan sebuah metode prediksi model terhadap infeksi hepatitis B dimana dalam penelitian ini, virus hepatitis diprediksi

dengan menggunakan klasifikasi Jaringan Saraf Tiruan (*Artificial Neural Network/ANN*), sedangkan algoritma *genetic algorithm* digunakan sebagai alat untuk memilih fitur-fitur yang berperan dalam terjadinya infeksi virus hepatitis B (jenis kelamin, hilangnya nafsu makan, mual dan muntah, kulit dan mata menguning, nyeri perut, serta nyeri pada otot dan persendian). Hasil prediksi menunjukkan kinerja yang dapat diterima, yang diharapkan dapat mengurangi insiden wabah di masa depan dan membantu respons cepat para ahli medis. Penelitian ini menggunakan sebanyak 900 data medis. Hasil penelitian menunjukkan bahwa model tersebut mencapai akurasi sebesar 66,30%, spesifisitas 66,33%, dan sensitivitas 77,53%.

Pada Artikel Wang et al. [6], mengusulkan satu model prediksi berbasis jaringan saraf graf dengan mengintegrasikan algoritma pembelajaran terawasi yang efisien, yaitu implementasi yang sangat baik dari strategi *gradient boosting*, GBDT. Dengan menggabungkan 12 set data klasifikasi biner yang optimal, satu model prediksi multi-target dibangun. Untuk mengevaluasi kinerja model *ensemble* prediksi multi-target kami, lima set data eksternal dibangun untuk evaluasi prediksi, dan semuanya mencapai PPV dan TPR yang memuaskan, yang berarti prediksi relatif akurat dari target potensial pada HIV-1 dan HBV. Dalam penelitian Marino et al [7], *Gated Graph Neural Network* dibahas sebagai varian dari *Graph Neural Network*. Dimana mekanisme gerbang atau *gate* diterapkan. Gerbang pada GGNN adalah sebuah metode yang mirip dengan *recurrent neural network*, *gate* tersebut disebut dengan *Gated Recurrent Unit*. Pada penelitian ini, ditekankan bahwa GGNN dapat memiliki stabilitas performa yang cukup dinamis.

Hepatitis

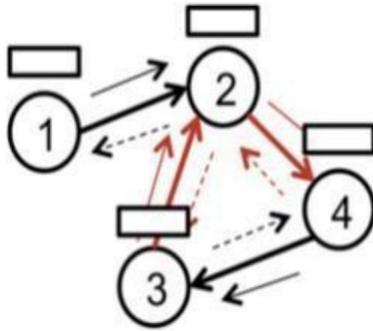
Hepatitis B virus merupakan sebuah infeksi pada hati yang dapat menyebabkan penyakit akut dan kronis [8], dimana Hepatitis adalah anggota famili Hepadnavirus yang dapat menyebabkan beberapa gejala samping yang cukup berbahaya terhadap tubuh manusia seperti peradangan hati yang akut atau menahun. Hepatitis tidak menyebar melalui bersin, batuk, maupun melewati makanan melainkan Hepatitis dapat tersebar melalui seseorang yang terinfeksi, akan tetapi seseorang tersebut tidak tahu akan ineksinya karena tidak merasakan atau terlihat sakit. Hepatitis dapat menyebar melalui hubungan kontak dengan darah seseorang yang terjangkit dengan Hepatitis, seperti halnya hubungan seks, menggunakan barang yang terkontaminasi seperti sikat gigi, silet, dan barang-barang medis.

Graph Neural Network

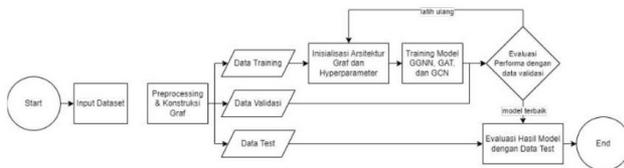
Graph neural Network (GNN) adalah sebuah arsitektur jaringan saraf yang didesain untuk bekerja dalam data yang berstruktur graf. Proses utama dalam GNN adalah melakukan penyebaran informasi melalui graf dengan mengumpulkan dan memperbaharui representasi simpul berdasarkan informasi graf terdekat [9].

Gated Graph Neural Network

Gated Graph Neural Network (GGNN), adalah sebuah bentuk lanjutan dari model *Graph Neural Network* (GNN), dimana diimplementasi adaptasi sehingga model sesuai dengan *non-sequential output*. Perubahan model GGNN yang terbesar dapat terlihat dari penggunaan *Gated Recurrent Unit* [10]. Pada *Gated Graph Neural Network*, mekanisme *gate* digunakan untuk mengatur aliran informasi dari *node* tetangga saat memperbarui representasi (state) *node*. GGNN menggunakan *Gated Recurrent Unit* (GRU) untuk secara dinamis mengendalikan bagaimana informasi dari iterasi sebelumnya dan informasi baru yang datang dari tetangga harus digabungkan. GGNN pertama-tama menghitung pesan dari tetangga menggunakan matriks bobot yang mengonversi state tetangga. Kemudian, pesan-pesan ini digabungkan untuk



Gambar 1. Bentuk ilustrasi Gated Graph Neural Network.



Gambar 2. Alur Diagram Penelitian.

membentuk satu vektor pesan yang merupakan agregasi dari informasi yang datang dari tetangga-tetangga *node*. Selanjutnya, GRU menggunakan vektor pesan ini bersama dengan state *node* sebelumnya untuk menentukan seberapa banyak informasi baru yang harus diterima dan seberapa banyak informasi lama yang harus dipertahankan. Mekanisme *gate* dalam GRU berfungsi seperti sebuah filter yang memungkinkan model untuk mempelajari representasi yang lebih kaya dan dinamis dengan menyesuaikan kontribusi dari informasi yang datang pada setiap iterasi. Dengan demikian, GGNN mampu mempertahankan konteks sejarah sekaligus menyesuaikan informasi baru yang datang, memberikan model kemampuan untuk memproses informasi pada jaringan graf secara lebih efektif dan akurat 1.

Metrik Evaluasi

Dari beberapa penelitian sebelumnya ada beberapa metrik evaluasi untuk menghitung akurasi yang dihasilkan model prediksi diagnosis Hepatitis, akan tetapi penggunaan *Accuracy* menjadi pilihan yang optimal dalam kasus ini, dikarenakan metrik ini berguna untuk mengukur sejauh mana model GGNN mampu memprediksi dengan benar kasus Hepatitis secara keseluruhan. Metrik *Precision* dan *Recall* juga cukup penting dalam konteks prediksi Hepatitis dimana metrik tersebut berguna untuk mengevaluasi seberapa baik model rancangan dalam mengidentifikasi kasus positif yang sebenarnya(presisi), dan seberapa baik dalam menemukan semua kasus positif yang sebenarnya(recall) [9].

Metodologi Penelitian

Pada penelitian ini tela dilakukan proses percobaan eksperimen sesuai pada alur diagram pada Gambar 2. Data yang digunakan akan dilakukan analisis dan preprocessing, yang kemudian dibangun kedalam bentuk graf. Graf tersebut kemudian dilakukan *splitting* berdasarkan jumlah *node* didalamnya, data dibagi menjadi data training data, validation data, dan test data. Pelatihan dilakukan pada data train dan di validasi dengan validation data, hasil model yang sudah dilakukan pelatihan kemudian dilakukan prediksi terhadap test data.

Age	Sex	Steroid	Antiviral	Fatigue	Malaise	Anorexia	Liver Big	Liver Firm	Spleen Palpable	Spiders	Ascites	Varices	Bilirubin	Alkaline Phosphatase	Albumin	Prothrombin Time	Histology	Class	
30	2	1	2	2	2	2	2	1	2	2	2	2	1	1	1	1	1	1	1
50	1	1	2	1	2	2	2	1	2	2	2	2	0	0	135.0	42.0	3.5	NaN	1
75	1	2	2	1	2	2	2	2	2	2	2	2	0	0	94.0	32.0	4.8	NaN	1
31	1	NaN	1	2	2	2	2	2	2	2	2	2	0	0	46.0	32.0	4.8	NaN	1
34	1	2	2	2	2	2	2	2	2	2	2	2	1	1	NaN	200.0	4.8	NaN	1
46	1	2	2	1	1	1	1	2	2	2	1	1	1	7.8	NaN	242.0	5.3	NaN	2
44	1	2	2	1	2	2	2	1	2	2	2	2	0	0	126.0	142.0	4.3	NaN	2
61	1	1	2	1	1	1	1	1	1	2	2	2	0	0	75.0	30.0	4.1	NaN	2
53	2	1	2	1	2	2	2	2	2	2	2	2	1	1	81.0	16.0	4.1	NaN	2
45	1	2	2	1	2	2	2	2	2	2	2	2	1	1	160.0	16.0	5.1	NaN	2

Gambar 3. Dataset Hepatitis UCI Machine Learning Repository.

Dataset

Dataset yang digunakan diambil dari UCI *Machine Learning Repository*, dimana didalamnya terdapat 19 fitur dari pasien hepatitis, dan labelnya 3.

Exploratory Data Analysis & Preprocessing

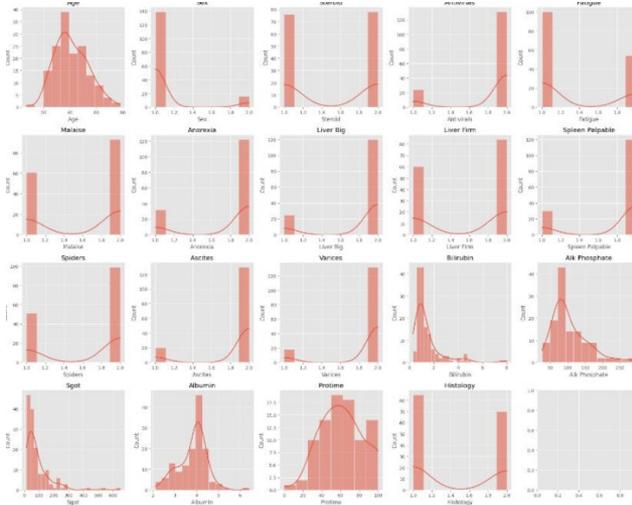
Pada dataset UCI terdapat 155 sampel, dimana masing masing sample memiliki 19 atribut. Atribut tersebut adalah sebagai berikut:

1. *Age* : terdapat jangkauan umur dari 20 sampai 72 tahun.
2. *Sex* : pria atau wanita, di representasikan dengan masing-masing 1 dan 2.
3. *Steroid* : nilai 1 untuk tidak, dan 2 untuk ya.
4. *Antiviral* : nilai 1 untuk tidak, dan 2 untuk ya.
5. *Fatigue* : nilai 1 untuk tidak, dan 2 untuk ya.
6. *Malaise* : nilai 1 untuk tidak, dan 2 untuk ya.
7. *Anorexia* : nilai 1 untuk tidak, dan 2 untuk ya.
8. *Liver Big* : nilai 1 untuk tidak, dan 2 untuk ya.
9. *Liver Firm* : nilai 1 untuk tidak, dan 2 untuk ya.
10. *Spleen Palpable* : nilai 1 untuk tidak, dan 2 untuk ya.
11. *Spiders* : tufa kapiler merah pada kulit yang pucat saat ditekan. Diwakili dengan 1 dan 2.
12. *Ascites* : penumpukan cairan di rongga perut, diwakili oleh angka 1 atau 2.
13. *Varices* : pembuluh darah yang melebar, diwakili oleh angka 1 atau 2.
14. *Bilirubin* : pigmen empedu yang dibersihkan dari darah oleh hati
15. *Alkaline* : protein yang ditemukan di membran sel saluran empedu
16. *Aspartate* : enzim yang mengkatalisis transformasi protein dalam hepatosit
17. *Albumin* : protein dalam serum yang mengangkut zat seperti obat-obatan dan mencegah kebocoran cairan ke jaringan sekitarnya.
18. *Pro-time* : waktu protrombin dalam serum
19. *Histology* : nilai 1 untuk tidak, dan 2 untuk ya.
20. *Class* : nilai 1 untuk terdiagnosis hepatitis, dan 2 untuk tidak terdiagnosis.

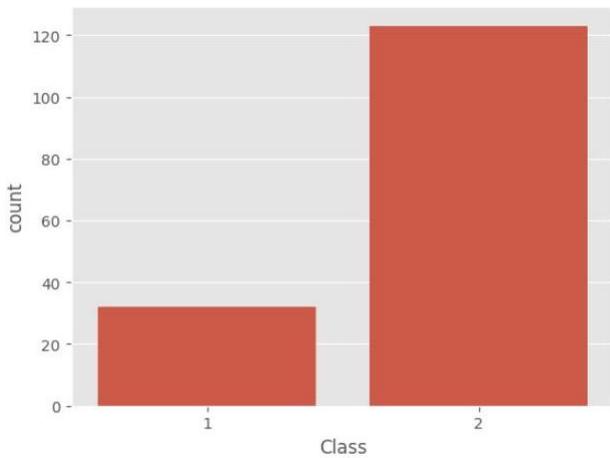
Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not

Pada Gambar 4. Dapat dataset *sex*(jenis kelamin), *steroid*, *antivirals*, *fatigue*, *malaise*, *anorexia*, *liver big*, *liver firm*, *spleen palpable*, *spiders*, *ascites*, *varices* dan *histology* memiliki nilai berupa nilai biner sehingga distribusi hanya terlihat pada dua bar yang berbeda. Data yang digunakan pada penelitian ini memiliki beberapa nilai yang hilang, karena nilai yang hilang pada data juga menjadi suatu permasalahan yang ada, dan absensi dari sebuah nilai tersebut dapat berpengaruh terhadap performansi dari model graf yang akan dilatih, oleh karena itu untuk mengatasi nilai- nilai tersebut dilakukan lah imputing data dengan cara *iterative imputing*. *Iterative imputer* bekerja dengan menggunakan kolom nilai yang hilang sebagai target dan kolom lainnya sebagai fitur, dan melakukan pelatihan data tersebut dengan model regresi.

Selain itu, terdapat juga ketidak seimbangan target prediksi pada dataset, ketidak seimbangan ini atau dapat disebut *imbalanced data*



Gambar 4. Diagram Chart Distribusi atribut.



Gambar 5. Label prediksi imbalance.



Gambar 6. Korelasi Fitur terhadap Label.

adalah sebuah kondisi dimana kelas-kelas pada target memiliki penyebaran yang tidak merata. Seperti halnya pada kelas label pada Gambar 5. Ketidak seimbangan ini dapat menyebabkan model klasifikasi menurun, akurasi yang rendah dan *overfitting* [11]. Oleh karena itu dilakukan penyeimbangan data untuk meningkatkan performa model. Oleh karena itu dilakukan metode penyeimbangan dengan cara *oversampling* pada data minoritas, metode *oversampling* yang digunakan pada data ini adalah dengan cara menambah minoritas secara acak dan menyesuaikan sebanyak 1:1

Setelah melakukan *imputing* dan *oversampling*, dilakukan korelasi matriks fitur terhadap target prediksi, pada gambar 6. terdapat beberapa hubungan fitur yang memiliki nilai negatif, yaitu *age*, *anti-virals*, *liverbig*, *bilirubin*, *alk phosphate*, dan *histology*. Sedangkan *sex*, *steroid*, *fatigue*, *malaise*, *anorexia*, *liver firm*, *spleen palpable*,

Table 1. *Hyperparameter* Pelatihan

<i>Hyperparameter</i>	Nilai
Epoch	200; 400; 600
Dropout	0.1; 0.2; 0.3

spiders, *ascites*, *varieces*, albumin dan *protime* memiliki pengaruh terhadap target secara positif, dengan Albumin yang memiliki pengaruh terbesar sebesar 0.61. Nilai korelasi 0.61 yang ditunjukkan untuk variabel Albumin terhadap *Class* menunjukkan adanya korelasi positif yang cukup kuat antara kadar albumin dalam dataset terhadap kelas label, dimana korelasi nilai albumin dapat mempengaruhi probabilitas *class* label sebanyak 61% dari nilai albumin. Begitu juga sebaliknya pada fitur yang nilai korelasinya adalah negatif, maka pengaruh probabilitas terhadap label *class* akan berbalik menurun.

Model Pelatihan

Data dari graf sebelumnya yang sudah dibuat, dilakukan data *splitting* dengan rasio sebesar 10% data validasi, 20% data tes, dan 70% *data training*. Pada *train model* GGNN ini dilakukan perbandingan antara beberapa metode *Graph Neural Network* yang lain, diantaranya adalah *Graph Attention Network* (GAT) dan *Graph Convolutional Network* (GCN). Ini dilakukan dengan tujuan untuk melakukan evaluasi hasil model terhadap metode yang lain. Graf data yang diinput kedalam model digunakan sebagai representasi numerik, dengan parameter *hidden layer* sebanyak 32 memungkinkan model untuk memproses data yang lebih kompleks, ini karena dengan lebih banyak *hidden layer* dapat memungkinkan data graf padat atau *dense* yang sebelumnya sudah dibuat, akan membuat model prediksi menjadi lebih baik. Fungsi *Rectified Linear Unit* (ReLU) diterapkan disetiap lapisan, ini dilakukan untuk memberikan representasi nonlinear dari fitur *node*. Kemudian dilakukan juga *Dropout* dimana berfungsi untuk mencegah *overfitting*, dengan cara menghapus secara acak bagian *network* pada saat pelatihan. Dan terakhir, *optimizer* Adam digunakan pada *run_experiment* untuk memperbarui bobot model selama pelatihan dengan tujuan untuk meminimalkan fungsi *loss*. Adam seringkali konvergen lebih cepat daripada optimisasi lain seperti *Stochastic Gradient Descent* (SGD) pada banyak masalah. Terutama pada inialisasi parameter yang kurang baik, Adam dapat merupakan *optimizer* yang lebih kokoh. Dan setelah semua tahap dilakukan, model akan melakukan pelatihan terhadap data secara ulang, *Training Loop*. Dengan adanya iterasi dari data graf, dan melakukan *Forward message*, untuk memperbarui *node*. Pada *looping* ini juga dihitung secara langsung hasil perbandingan data latih dengan data valid, dimana fungsi *Cross Entropy Loss* digunakan.

Skenario Percobaan

Pada uji percobaan model, dilakukan percobaan berdasarkan *hyperparameter tuning*, dimana parameter ini digunakan untuk mendapatkan parameter terbaik dalam proses prediksi model. Percobaan ini menggunakan parameter *dropout* dan *epoch*, sebagai bentuk evaluasinya. Pada proses percobaan akan digunakan *Cross Entropy Loss* sebagai fungsi perhitungan model terbaik dalam pengujian.

Hasil dan Pembahasan

Skenario percobaan dilakukan dengan *hyperparameter tuning* yang sudah ditentukan sebelumnya. Tabel 1. Merupakan tabel hasil percobaan dimana terdiri dari 3 kolom utama, yaitu *epoch*, *dropout*, dan kolom hasil *Cross Entropy Loss* terbaik dari setiap kategori *epoch* dan *dropout*. Setelah dilakukan *training model* dengan menggunakan parameter yang sudah ditentukan, hasil *loss* dari GCN selalu meningkat pada saat penggunaan *dropout* sebesar 0.3, ini bisa disebabkan karena

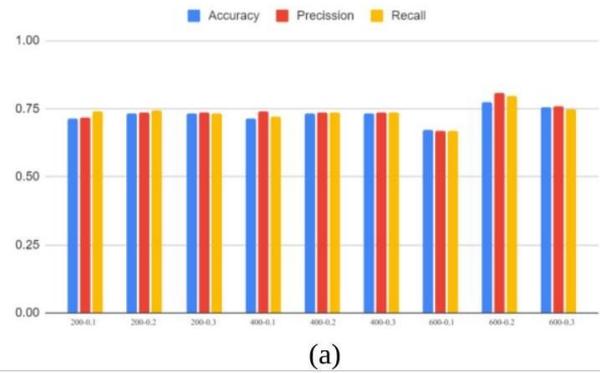
Table 2. Hasil *Loss* Pelatihan Model berdasarkan *Hyperparameter*

Epoch	Dropout	Data Validasi		
		GCN	GGNN	GAT
200	0.1	0.5858	0.1040	0.4680
	0.2	0.5650	0.0864	0.5132
	0.3	0.6467	0.1203	0.6300
400	0.1	0.3746	0.0123	0.3426
	0.2	0.6879	0.1968	0.3343
	0.3	0.8186	0.0667	0.4359
600	0.1	0.4656	0.1864	0.2758
	0.2	0.4871	0.0642	0.3969
	0.3	0.9787	0.2906	0.3253

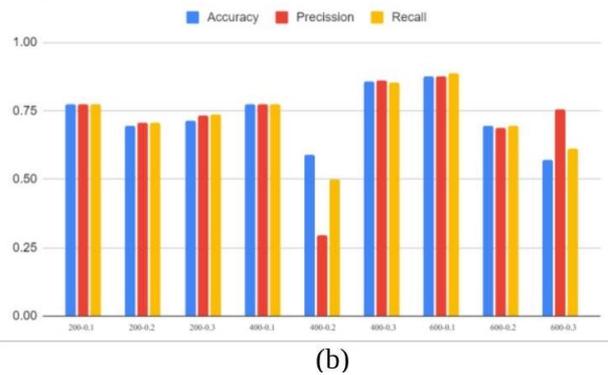
GCN memiliki sistem agregasi yang cukup sederhana dengan cara langsung mengkalikan bobot dengan fungsi aktifasinya. Sedangkan pada perhitungan *loss* pada GAT, *loss* yang didapat memiliki nilai yang lebih baik dibanding oleh GCN, ini dapat disebabkan oleh metode agregasi antar tetangga model yang dimana GAT menggunakan *attention weight* sebagai penentu sebagaimana banyak sebuah *node* akan bergantung pada *node* lain. Sehingga perpindahan informasi pada graf dapat berjalan lebih baik. Dan di model prediksi utama yang di analisis pada penelitian ini, GGNN mendapat nilai *loss* yang rendah di setiap kategori dibanding GAT dan GCN. Dengan fungsi GRU sebagai metode propagasi informasi antar tetangganya, ini dapat dikatakan bahwa gerbang pada GGNN atau GRU memiliki peran besar dalam menentukan sebagaimana banyak informasi yang perlu di perbaharui. Dari hasil pelatihan model, dapat dilihat bahwa nilai *loss* pada *dropout* sebesar 0.3 rata-rata memiliki nilai yang lebih besar dibanding *dropout* yang lain. Ini dapat disebabkan oleh dataset yang tergolong kecil terhadap rasio *dropout* yang digunakan, dimana dataset UCI memiliki *instance* sebanyak 155. Melakukan *dropout* sebanyak 30% dari total seluruh data dapat menyebabkan model prediksi berjalan secara tidak maksimal. Dibandingkan dengan melakukan *dropout* sebesar 10% dari dataset, pelatihan model dapat mendapatkan nilai *loss* yang cukup rendah. Oleh karena itu, dapat dikatakan rasio ini memungkinkan model yang dilatih untuk melakukan lebih banyak percobaan pada lebih banyak data dan menghasilkan prediksi yang lebih baik.

Hasil terbaik dari percobaan ini, dapat ditentukan bahwa GCN memiliki *Loss* terendah pada *epoch* ke-400 dengan *dropout* sebesar 0.1. Kemudian GAT pada kategori *epoch* ke-600 dengan *dropout* sebesar 0.2, dan terakhir GGNN dapat menghasilkan *loss* terkecil dari seluruh model sebesar 0.0864 pada *epoch* ke-400 dengan *dropout* sebesar 0.1 2. Pada Skenario berikutnya, hasil pelatihan model pada setiap kategori *epoch* dan *dropout* digunakan model dengan *loss* terbaiknya dan dilakukan evaluasinya terhadap data test. Evaluasi ini dilakukan dengan metrik *Accuracy*, *Precision*, dan *recall* 3. Berdasarkan hasil prediksi yang dilakukan terhadap data tes, GGNN menggapai skor evaluasi terbaik pada *epoch* ke-200 dengan *dropout* sebesar 0.1, dimana skor akurasi yang didapat adalah 93% *precision* sebesar 93%, dan *recall* sebesar 94%. Pada Gambar 8. GGNN mendominasi prediksi dari data ini dengan rata rata akurasi sebesar 86% dari seluruh kategori *Hyperparameter*. Sedangkan GCN mendapat rata-rata sebesar 73% dan GAT sebesar 72%. Pada hasil analisis ini dapat dibandingkan bahwa GGNN memiliki akurasi yang sedikit lebih baik 7, dibanding dengan model *Xgboost* yang mendapatkan akurasi prediksi sebesar 92% Obaido et al. [2]. Akan tetapi perlu diketahui pada penelitian ini mengadaptasi model graf sebagai bentuk representasi data hepatitis, sebagaimana tidak digunakan pada penelitian *Xgboost*. Sehingga perlu

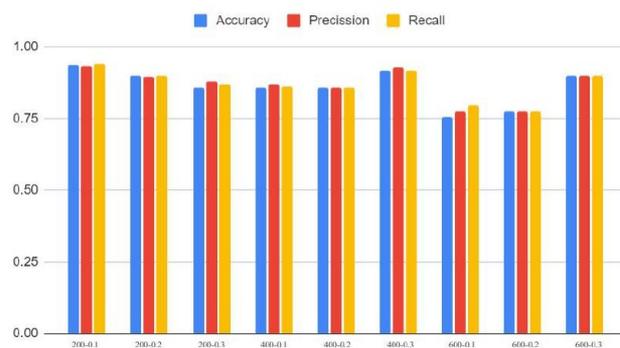
Evaluasi GCN



Evaluasi GAT

**Gambar 7.** Chart Hasil Prediksi terhadap data tes (a) Hasil Evaluasi GCN (b) Hasil Evaluasi GAT.

Evaluasi GGNN

**Gambar 8.** Chart Evaluasi GGNN terhadap Data Tes.

diketahui bahwa penggunaan graf dengan model GGNN masih dapat dilakukan, dengan kinerja performa yang masih berada pada akurasi sebesar 93%.

Kesimpulan

Pada percobaan training model GGNN mendapatkan nilai *loss* terkecil diantara GCN dan juga GAT, dengan itu hasil *loss* ini menjadi acuan terhadap evaluasi model terhadap data tes, alhasil prediksi diagnosis Hepatitis dengan menggunakan GGNN, mendapatkan nilai terbaik pada parameter *epoch* sebanyak 200 kali dengan *dropout* sebesar 0.1.

Table 3. Hasil Evaluasi Model terhadap Data Tes

Evaluasi model terhadap data tes										
Epoch	Dropout	GCN			GGNN			GAT		
		Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
200	0.1	0.7143	0.7175	0.7399	0.9388	0.9328	0.9404	0.7755	0.7759	0.7750
	0.2	0.7347	0.7367	0.7448	0.8980	0.8963	0.8990	0.6939	0.7083	0.7083
	0.3	0.7347	0.7381	0.7333	0.8571	0.8793	0.8704	0.7143	0.7321	0.7353
400	0.1	0.7143	0.7392	0.7232	0.8571	0.8690	0.8629	0.7755	0.7759	0.7750
	0.2	0.7347	0.7358	0.7382	0.8571	0.8575	0.8575	0.5918	0.2959	0.5000
	0.3	0.7347	0.7358	0.7382	0.9184	0.9310	0.9167	0.8571	0.8631	0.8558
600	0.1	0.6735	0.6700	0.6700	0.7551	0.7736	0.7987	0.8776	0.8758	0.8888
	0.2	0.7755	0.8070	0.7976	0.7755	0.7759	0.7750	0.6939	0.6890	0.6948
	0.3	0.7551	0.7614	0.7492	0.8980	0.8988	0.8975	0.5714	0.7558	0.6111

Selain itu, GGNN pada evaluasi data tes disemua kategori epoch mendapatkan evaluasi dengan akurasi rata-rata 86%. Dibanding dengan kedua metode Graph Neural Network lainnya yang juga diuji, GGNN menampilkan performa yang lebih baik diantaranya.

Daftar Pustaka

- World Health Organization. Hepatitis; 2024. Accessed 4 September 2024. Available from: <https://www.who.int/health-topics/hepatitis>.
- Obaido G, Ogbuokiri B, Swart TG, Ayawei N, Kasongo SM, Aruleba K, et al. An Interpretable Machine Learning Approach for Hepatitis B Diagnosis. *Applied Sciences*. 2022;12(21):11127. Available from: <https://doi.org/10.3390/app122111127>.
- Ibrahim D, Ahmadu AS, Malgwi YM, Ahmad BM. The Prediction of Hepatitis B Virus (HBV) Using Artificial Neural Network (ANN) and Genetic Algorithm (GA). *Computer Science & IT Research Journal*. 2021;2(1):16-32.
- Chen F, Wang YC, Wang B, Kuo CCJ. Graph Representation Learning: A Survey. *APSIPA Transactions on Signal and Information Processing*. 2020;9:e15.
- Sun Z, Yin H, Chen H, Chen T, Cui L, Yang F. Disease Prediction via Graph Neural Networks. *IEEE Journal of Biomedical and Health Informatics*. 2021;25(3):818-26.
- Wang Y, Li Y, Chen X, Zhao L. HIV-1/HBV Coinfection Accurate Multitarget Prediction Using a Graph Neural Network-Based Ensemble Predicting Model. *International Journal of Molecular Sciences*. 2023;24(8):7139. Available from: <https://doi.org/10.3390/ijms24087139>.
- Marino A, Pacchierotti C, Giordano PR. On the Stability of Gated Graph Neural Networks. 2023. Hal-04110671v1.
- Pattyn J, Hendrickx G, Vorsters A, Van Damme P. Hepatitis B Vaccines. *The Journal of Infectious Diseases*. 2021;224(12 Suppl 2):S343-51. Available from: <https://doi.org/10.1093/infdis/jiaa668>.
- Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: A review of methods and applications. *AI Open*. 2020;1:57-81.
- Ruiz L, Gama F, Ribeiro A. Gated Graph Recurrent Neural Networks. *IEEE Transactions on Signal Processing*. 2020;68:6303-18.
- Ghorbani M, Kazi A, Baghshah MS, Rabiee HR, Navab N. RA-GCN: Graph Convolutional Network for Disease Prediction Problems with Imbalanced Data. *Medical Image Analysis*. 2022;75:102272.